

## Document 4: Data & AI Strategy

**Document ID:** LKP-DAS-002

**Document Version:** 2.0 (Hybrid AI MVP)

**Status:** Approved for Development

**Date:** 22-08-2025

**Owner:** AI/Prompt Engineer & Research Specialist

**Document Purpose:** To define the core AI conversational logic, the "Master Prompts" for the LLM, and the structure of the static data that powers the "NextLeap" MVP. This document is the blueprint for the application's intelligence and knowledge base.

**Intended Audience:** AI/Prompt Engineer, Backend Developer, Research Specialist, Team Lead.

---

### 1.0 The AI Strategy: The "Hybrid Brain"

Our AI strategy is a unique hybrid model designed to leverage the best of both worlds: the creative power of LLMs and the reliability of structured data.

1. **The Dynamic Brain (The Conversationalist):** This is the **Google Gemini API**. Its job is to conduct a natural, empathetic conversation to understand the user's core profile (e.g., Analytical, Creative). It handles all the complex permutations of user interests.
2. **The Structured Brain (The Fact-Checker):** This is our set of manually curated **CSV files**. Its job is to provide the final, factually accurate, and reliable data for the roadmap (colleges, careers).

The Dynamic Brain finds the "direction," and the Structured Brain provides the "map."

---

### 2.0 The Conversational Engine: Prompt Engineering

Our core IP is our "Master Prompt." This is the detailed instruction set we provide to the Gemini API to guide its behavior.

codeText

```
# Master Prompt for NextLeap AI
```

```
**Persona:**
```

```
You are 'Cogni', a friendly, encouraging, and expert career counselor for Indian students who have just completed 10th grade. Your primary language is Hinglish. You are modern, use emojis where appropriate (like 🚀, 🧠, 📚), and make the student feel relaxed and understood.
```

```
**Primary Goal:**
```

Your goal **is to** have a short, **3 to 4** step conversation **to** deeply understand the student's core interests and passion. Based on the conversation, you must determine if their primary inclination is towards **\*\*Analytical\*\***, **\*\*Creative\*\***, or **\*\*Managerial\*\*** fields.

**\*\*Conversation Rules:\*\***

1. **\*\*Always ask a Multiple-Choice Question.\*\***
2. **\*\*Provide 2 to 4 clear, concise, and clickable options.\*\***
3. **\*\*Your next question MUST be a logical follow-up to the user's previous answer.\*\*** **Show** that you **are** listening.
4. **\*\*Never ask about marks or academic scores.\*\*** The focus **is** purely **on** interest **and** passion.
5. **\*\*After 3-4 questions, you must have enough information to make a logical recommendation.\*\***
6. **\*\*The conversation starts after the user's first input (e.g., "After 10th").\*\*** Your first question should be broad, like about their favorite subject.

**\*\*Output Format:\*\***

Your response **MUST ALWAYS** be a valid JSON object. Do not add any text before or after the JSON object.

\* **\*\*If the conversation is ongoing, use this format:\*\***

```
{
  "type": "question",
  "payload": {
    "next_question": "Your generated question here...",
    "options": ["Option A", "Option B", "Option C"]
  }
}
```

\* **\*\*When you have enough information to decide, use this format:\*\***

```
{
  "type": "recommendation",
  "payload": {
    "logical_recommendation": "The final career cluster you have decided (e.g., Engineering)",
    "recommendation_reason": "A brief, one-line reason for your choice."
  }
}
```

The logic for reacting to the Happiness Slider score will be handled in our **FastAPI backend**, not by the LLM.

- **If score is 4 or 5:** The backend confirms the `logical_recommendation` and proceeds to generate the roadmap.
- **If score is 1 or 2:** The backend will identify a pre-defined "alternative" path. For example, if the logical path was `Engineering (Analytical)`, the alternative could be `Data Science (Analytical + Creative)`.
- **If score is 3:** The backend will proceed with the logical recommendation but add a note in the final roadmap about exploring alternatives.

---

### 3.0 The Knowledge Base: Static Data Strategy

This data is the factual foundation of our final roadmap. It must be well-researched and structured.

- **Colleges & Exams:** Information will be curated based on the official websites of **NIRF**, **NTA**, and the **Higher Education Department of J&K**.
- **Careers & Skills:** Information will be curated based on trends from professional platforms like **LinkedIn** and **Naukri.com**.

The following CSV files must be created and placed in the `/data` directory.

#### 1. `careers.csv`

- **Purpose:** To provide details for the final recommended career path. The `Career_Cluster` is the key for filtering.
- **Structure:**

Career_ID	Career_Cluster	Career_Name	Description_en	Description_hi	Top_Skill_1	Top_Skill_2
C01	Engineering	Engineering (CSE)	...	...	Python	Data Structures
C02	Management	Management (BBA)	...	...	Communication	Analytics
C03	Design	Design (UI/UX)	...	...	Figma	User Research
C04	Data Science	Data Science	...	...	Statistics	Python

#### 2. `colleges.csv`

- **Purpose:** To provide college recommendations. The Career\_Cluster column is the primary filter. A is\_gov\_jk flag is added for the specific problem statement.
- **Structure:**

College_ID	College_Name	City	Primary_Exam_for_UG	Career_Cluster	is_gov_jk
101	IIT Jammu	Jammu	JEE Advanced	Engineering	No
102	NIT Srinagar	Srinagar	JEE Main	Engineering	No
103	GDC Udampur	Udhampur	Merit-Based	Management	Yes

(At least 15-20 entries with a focus on J&K Government Colleges)

### 3. checklists.csv (For the Printable Checklist)

- **Purpose:** To store the actionable tasks for the downloadable checklist.
- **Structure:**

Checklist_ID	Career_Cluster	Phase	Task_en	Task_hi
CL01	Engineering	Foundation	Finalize Computer Science as the 5th subject.	कंप्यूटर साइंस को 5वें विषय के रूप में फाइनल करें।
CL02	Engineering	Foundation	Talk to at least one senior preparing for engineering.	इंजीनियरिंग की तैयारी कर रहे कम से कम एक सीनियर से बात करें।
CL03	Management	Foundation	Participate in a school debate or event.	स्कूल की किसी बहस या कार्यक्रम में भाग लें।